

**REMARKS**

By this amendment, claims 1-16 and 21-28 are pending in the present application, of which claims 1-3 and 27-28 are currently amended. Claims 17-20 have been previously canceled without prejudice or disclaimer. No new matter is introduced.

The Office Action dated September 13, 2010:

(1) rejected claims 1 and 3 under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as the invention;

(2) rejected claims 1, 3-12 and 21-27 under 35 U.S.C. § 103(a) as being unpatentable over *Smith et al.* (Disambiguating Geographic Names in a Historical Digital Library) in view of *Wacholder et al.* (Disambiguation of Proper Names in Text) and *Bagga et al.* (Entity-Based Cross-Document Coreference Using the Vector Space Model);

(3) rejected claims 2 and 28 under 35 U.S.C. § 103(a) as being unpatentable over *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Frank et al.* (WO 01/63479 A1);

(4) rejected claim 13 under 35 U.S.C. § 103(a) as being unpatentable over *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Naughton* (US 6,240,425); and

(5) indicated that claims 14-16 are objected to as being dependent upon a rejected base claim, but would be allowable if rewritten in independent form including all of the limitations of the base claim and any intervening claims.

**A. 35 U.S.C. § 112, Second Paragraph, Rejection of Claims 1 and 3**

Claims 1 and 3 stand rejected under 35 U.S.C. § 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which Applicants regard as the invention (Office Action, Pp. 2-3). To advance prosecution,

Applicants have amended claims 1 and 3 to clarify “the references to other items in the claims” that the Examiner alleges as being unclear. Additionally, even though not addressed by the Examiner, claims 27 and 28 contained similar references – so Applicants have made similar amendments to claims 27 and 28. Accordingly, withdrawal of the rejection is respectfully requested.

**B. 35 U.S.C. § 103(a) Rejection of Claims 1, 3-12 and 21-27 Over *Smith* In View Of *Wacholder* and *Bagga***

---

Applicants respectfully traverse the 35 U.S.C. § 103(a) rejection of claims 1, 3-12 and 21-27 over *Smith* in view of *Wacholder* and *Bagga*, because all features of the claims are not disclosed by the applied art, either individually or in combination.

Specifically, independent claim 1 recites, *inter alia*, “determining a value for a confidence that the selected toponym is associated with the selected reading, wherein determining said value involves a mathematical summation over the plurality of documents in which geo-textual correlations were identified that involved the selected toponym and the selected reading.” Independent claim 27 similarly recites, *inter alia*, “determin[ing] a value for a confidence that the selected toponym is associated with the selected reading based, at least in part, on a mathematical summation over the plurality of documents in which geo-textual correlations were identified that involved the selected toponym and the selected reading.” Applicants submit, as presented below, that neither *Smith*, *Wacholder* or *Bagga* alone, nor the combination of *Smith* in view of *Wacholder* and *Bagga*, discloses or suggests such features.

With respect to claims 1 and 27, the Examiner acknowledges that the teachings of *Smith* and *Wacholder* lack the disclosure of the claimed element of “wherein determining said value involves a mathematical summation over the plurality of documents in which geo-textual

correlations were identified that involved that toponym-reading pair.” (*Office Action*, P. 5, LL. 14-16) Instead, the Examiner cites to *Bagga* for the alleged disclosure of this claimed element. (*Office Action*, P. 5, L. 17 to P. 6, L. 2) The relevant disclosure of *Bagga* cited by the Examiner is as follows:

The vector space model used for disambiguating entities across documents is the standard vector space model used widely in information retrieval (Salton 89). In this model, each summary extracted by the SentenceExtractor module is stored as a vector of terms. The terms in the vector are in their morphological root form and are filtered for stop-words (words that have no information content like *a*, *the*, *of*, *an*, ...). If  $S_1$  and  $S_2$  are the vectors for the two summaries extracted from documents  $D_1$  and  $D_2$ , then their similarity is computed as:

$$Sim(S_1, S_2) = \sum_{\text{common terms } t_j} w_{1j} \times w_{2j}$$

where  $t_j$  is a term present in both  $S_1$  and  $S_2$ ,  $w_{1j}$  is the weight of the term  $t_j$  in  $S_1$  and  $w_{2j}$  is the weight of  $t_j$  in  $S_2$ .

The weight of a term  $t_j$  in the vector  $S_i$  for a summary is given by:

$$w_{ij} = \frac{tf \times \log \frac{N}{df}}{\sqrt{s_{i1}^2 + s_{i2}^2 + \dots + s_{in}^2}}$$

where  $tf$  is the frequency of the term  $t_j$  in the summary,  $N$  is the total number of documents in the collection being examined, and  $df$  is the number of documents in the collection that the term  $t_j$  occurs in.  $\sqrt{s_{i1}^2 + s_{i2}^2 + \dots + s_{in}^2}$  is the *cosine normalization factor* and is equal to the Euclidean length of the vector  $S_i$ .

With reference to this disclosure, the Examiner alleges that the weight  $w_{ij}$  taught by *Bagga* can be characterized as a confidence value involving a mathematical summation over a plurality of documents. Specifically, the Examiner alleges as follows:

*Bagga* teaches a method of cross-document coreferencing when the same person, place, event is discussed in more than one text source (*Bagga*, Introduction). *Bagga* further discloses the formula for calculating the **score of a te[r]m**, i.e., the weight of a term in a vector of terms (*Bagga*, Page 81, Right Column-Lines 4-15) to determine the similarity of two document represented by extracted terms vectors. The **weight of a term t** as taught by *Bagga* is based on **df, which is the number of documents in the collection that the term t occurs in**. The variable **df** of documents over the collection of document **is a mathematical summation over the plurality of documents**, and within the collection, geographic textual correlations such as the occurrence of "Lancaster" that references to the occurrences of "Philadelphia" and "Harrisburg" are identified that involved the term "Philadelphia" and "Lancaster". (*Office Action*, P. 5, L. 17 to P. 6, L. 2)(*emphasis added*)

The Applicants respectfully and categorically disagree. First, the Applicants assert that although **df**, the number of documents containing a specific term  $t$ , in a strict mathematical sense is a (trivial) summation of 1's over a collection of documents, in a generic situation one of ordinary skill in statistical science would not characterize the number of documents **df** as a mathematical summation over a collection of documents. Nevertheless, the Examiner's assertion could have been mathematically true if the weight  $w_{ij}$  were a confidence value (or alternatively a "precision score" as referred to by *Bagga* in Section 7.2, for example) in associating a specific toponym with a specific reading. Examining the teaching of *Bagga*, it is evident, in view of the context of the term "weight" as used by *Bagga* and the formula of *Bagga*, that the weight that  $w_{ij}$  is neither (1) a confidence value (or precision score of any associating), nor (2) a confidence or score in associating toponym with a reading, for the following reasons:

1. As is readily apparent, a confidence value about a statement or an event is meant to indicate a belief (or equivalently reduces the **uncertainty**) that the statement is true or

that said event might happen. Accordingly, the higher the confidence value about a statement or an event, the more likely the statement is true or the event will happen. The weight or portion of a term in a text, however, can by no means be considered as a confidence number since such a number does not reasonably reduce uncertainty about any association. It should be noted that, the term “confidence” or “score” is a very common term in information retrieval art, and, the mere fact that *Bagga* uses the term “score” for several other quantities but does not refer to  $w_{ij}$  as a “score,” further shows that the characterization of  $w_{ij}$  as a score or confidence value is incorrect.

2. Furthermore, even assuming that  $w_{ij}$  as taught by *Bagga* could be characterized as a score (an assumption to which the Applicants by no means agree), as defined by *Bagga*, at best,  $w_{ij}$  would indicate a confidence or precision score in associating a **term  $t_j$  to a document  $D_i$** . Such an association, although useless in the context of both *Bagga*’s and the Applicants’ teachings, is completely and inherently different from the form of the association recited in claims 1 and 27 (i.e., the association of a **toponym with a reading of it**). One of ordinary skill in the art, therefore, would not and could not use a score function which gives a score to an association of a term to a document as a score function which scores an association of a toponym to a reading of it. In particular, it is not clear how, and by all means meaningless, to use the formula for  $w_{ij}$  in the context of associating a toponym to a reading of it, as recited in claims 1 and 27.

For at least the foregoing reasons, *Bagga* lacks the disclosure suggestion of the feature whereby the confidence value involves or is based on a “mathematical summation over the plurality of documents in which geo-textual correlations were identified that involved the

selected toponym and the selected reading,” as recited in claims 1 and 27, and thus fails to overcome the deficiencies of the combination of *Smith* and *Wacholder*.

Further, independent claim 10 recites, *inter alia*, “obtaining a pre-computed number for a value of a confidence that the toponym of the selected toponym-place pair refers to the place of the selected toponym-place pair, said pre-computed number derived from a statistical observation about a plurality of documents.” Applicants submit, as presented below, that neither *Smith*, *Wacholder* or *Bagga* alone, nor the combination of *Smith* in view of *Wacholder* and *Bagga*, discloses or suggests such features.

With respect to independent claim 10, the Examiner acknowledges that the teachings of *Smith* and *Wacholder* lack the disclosure of the claimed element of a “pre-computed number derived from a statistical observation about a large corpus of documents.” (*Office Action*, P. 9, Ll. 15-16) Instead, the Examiner cites also to *Bagga* for the alleged disclosure of this claimed element. (*Office Action*, P. 9, Ll. 17-24) Specifically, the Examiner alleges as follows:

*Bagga* teaches a method of cross-document coreferencing when the same person, place, event is discussed in more than one text source (*Bagga*, Introduction). **Bagga further discloses the formula for calculating the score of a t[e]rm, i.e., the weight** of a term in a vector of terms (*Bagga*, Page 81, Right Column-Lines 4-15) to determine the similarity of two document represented by extracted terms vectors. **The weight** of a term *t* as taught by *Bagga* is based on *df*, which is the number of documents in the collection that the term *t* occurs in. The variable *df* of documents over the collection of document is *pre-computed number derived from a statistical observation about a large corpus of documents.* (*Office Action*, P. 9, Ll. 17-24)

The Applicants respectfully disagree. Quite similarly to the previous argument, and contrary to the Examiner’s suggestion, for at least the following two reasons, the weight  $w_{ij}$  cannot be characterized or serve as a confidence value that the toponym of the selected toponym-place pair refers to the place of the selected toponym-place pair:

- 1 As is readily apparent, as stated above, a confidence value about a statement or an event is meant to indicate a belief (or equivalently reduces the **uncertainty**) that the statement is true or that said event might happen. Accordingly, the higher the confidence value about a statement or an event, the more likely the statement is true or the event will happen. The weight or portion of a term in a text, however, can by no means be considered as a confidence number since such a number does not reasonably reduce uncertainty about any association. It should be noted that, the term “confidence” or “score” is a very common term in information retrieval art, and, the mere fact that *Bagga* uses the term “score” for several other quantities but does not refer to  $w_{ij}$  as a “score,” further shows that the characterization of  $w_{ij}$  as a score or confidence value is incorrect.
- 2 Furthermore, even assuming that  $w_{ij}$  as taught by *Bagga* could be characterized as a score (an assumption to which the Applicants by no means agree), as defined by *Bagga*, at best,  $w_{ij}$  would indicate a confidence or precision score in associating a **term  $t_j$  to a document  $D_i$** . Such an association, although useless in the context of both *Bagga*’s and the Applicants’ teachings, is completely and inherently different from the form of the association recited in claim 10 (i.e., the association of the **toponym of a toponym-place pair to the place of the toponym-place pair**). One of ordinary skill in the art, therefore, would not and could not use a score function which gives a score to an association of a term to a document as a score function which scores an association of a toponym to place. In particular, it is not clear how, and by all means meaningless, to use the formula for  $w_{ij}$  in the context of associating a toponym to a place, as recited in claim 10.

For at least the foregoing reasons, *Bagga* lacks the disclosure suggestion of the feature of “obtaining a pre-computed number for a value of a confidence that the toponym of the selected toponym-place pair refers to the place of the selected toponym-place pair, said pre-computed number derived from a statistical observation about a plurality of documents,” as recited in claim 10, and thus fails to overcome the deficiencies of the combination of *Smith* and *Wacholder*.

Accordingly, for at least the foregoing reasons, neither *Smith*, *Wacholder* or *Bagga* alone, nor the combination of *Smith* in view of *Wacholder* and *Bagga*, render independent claims 1, 10 and 27, or claims 3-12 and 21-26 depending therefrom, obvious under 35 U.S.C. § 103.

**C. 35 U.S.C. § 103(a) Rejection of Claims 2 and 28 Over *Smith* In View of *Wacholder* and *Bagga*, and Further In View of *Frank***

---

Applicants respectfully traverse the 35 U.S.C. § 103(a) rejection of claims 2 and 28 over *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Frank*, because all features of the claims are not disclosed by the applied art, either individually or in combination.

Claims 2 and 28 depend from independent claims 1 and 27, respectively, and the Office Action applies the combination of *Smith* in view of *Wacholder* and *Bagga* to claims 2 and 28 on the same bases as with the § 103(a) rejection of their respective independent claims (addressed in Section B, above). Applicants incorporate herein the arguments presented above in Section B with respect to the application of *Smith* in view of *Wacholder* and *Bagga* to claims 2 and 28, accordingly. The Office Action cites to *Frank* for the alleged disclosure of the element of using the value for the confidence to rank documents according to their relevance to a search query, as recited in claims 2 and 28. (*Office Action*, P. 12) Applicants submit, however, that *Frank* lacks the disclosure or suggestion of the features whereby the confidence value involves or is based on a “mathematical summation over the plurality of documents in which geo-textual



correlations were identified that involved the selected toponym and the selected reading,” as recited in the independent claims 1 and 27, and thus fails to remedy the deficiencies of *Smith* in view of *Wacholder* and *Bagga*. Accordingly, for at least the foregoing reasons, neither *Smith*, *Wacholder*, *Bagga* and *Frank* alone, nor the cited combination of *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Frank*, render claims 2 and 28 obvious under 35 U.S.C. § 103.

**D. 35 U.S.C. § 103(a) Rejection of Claim 13 Over *Smith* In View of *Wacholder* and *Bagga*, and Further In View of *Naughton***

---

Applicants respectfully traverse the 35 U.S.C. § 103(a) rejection of claim 13 over *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Naughton*, because all features of the claims are not disclosed by the applied art, either individually or in combination.

Claim 13 depends from independent claim 1, and the Office Action applies the combination of *Smith* in view of *Wacholder* and *Bagga* to claim 13 on the same bases as with the § 103(a) rejection of its respective independent claims (addressed in Section B, above). Applicants incorporate herein the arguments presented above in Section B with respect to the application of *Smith* in view of *Wacholder* and *Bagga* to claim 13, accordingly. The Office Action cites to *Naughton* for the alleged disclosure of the element of “computing a geographical distance between the place associated with the identified toponym and the place referred to by the selected toponym-place pair,” as recited in claim 13. (*Office Action*, P. 13) Applicants submit, however, that *Naughton* lacks the disclosure or suggestion of the features whereby the confidence value involves or is based on a “mathematical summation over the plurality of documents in which geo-textual correlations were identified that involved the selected toponym and the selected reading,” as recited in the independent claim 1, and thus fails to remedy the deficiencies of *Smith* in view of *Wacholder* and *Bagga*. Accordingly, for at least the foregoing

reasons, neither *Smith*, *Wacholder*, *Bagga* and *Frank* alone, nor the cited combination of *Smith* in view of *Wacholder* and *Bagga*, and further in view of *Naughton*, render claim 13 obvious under 35 U.S.C. § 103.

**E. Conclusion**

Therefore, the present application, as amended, overcomes the objections and rejections of record and is in condition for allowance. Favorable consideration is respectfully requested. If any unresolved issues remain, it is respectfully requested that the Examiner telephone the undersigned attorney at (703) 519-9952 so that such issues may be resolved as expeditiously as possible.

To the extent necessary, a petition for an extension of time under 37 C.F.R. § 1.136 is hereby made. Please charge any shortage in fees due in connection with the filing of this paper, including extension of time fees, to Deposit Account 504213 and please credit any excess fees to such deposit account.

Respectfully Submitted,

DITTHAVONG MORI & STEINER, P.C.

December 13, 2010  
Date

/Phouphanomketh Ditthavong/  
Phouphanomketh Ditthavong  
Attorney/Agent for Applicant(s)  
Reg. No. 44658

Craig L. Plastrik  
Attorney/Agent for Applicant(s)  
Reg. No. 41254

918 Prince Street  
Alexandria, VA 22314  
Tel. (703) 519-9951  
Fax (703) 519-9958